The Rich the Poor and the Attractive: Global Class Bias

Anita Lilla Verő
Department of Computer Science
Cambridge University
Cambridge, United Kingdom
alv34@cam.ac.uk

Abstract

Twitter's saliency based cropping has been shown to be biased towards cropping lighter over darker skinned individuals and women over men. The question arises: Can class based bias be another harmful consequence of using this method? I tested the algorithm for images of *cheap* versus *expensive* objects, using the income labels as proxy in the dataset of Dollar Street. I found that the algorithm is biased towards *cheap* rooms and spaces and towards *expensive* objects. I deem that this behaviour could potentially cause several types of unintentional harm including: stereotyping, erasure, under-representation and reputational harm.

Keywords algorithmic bias · class bias · image cropping

1 Introduction

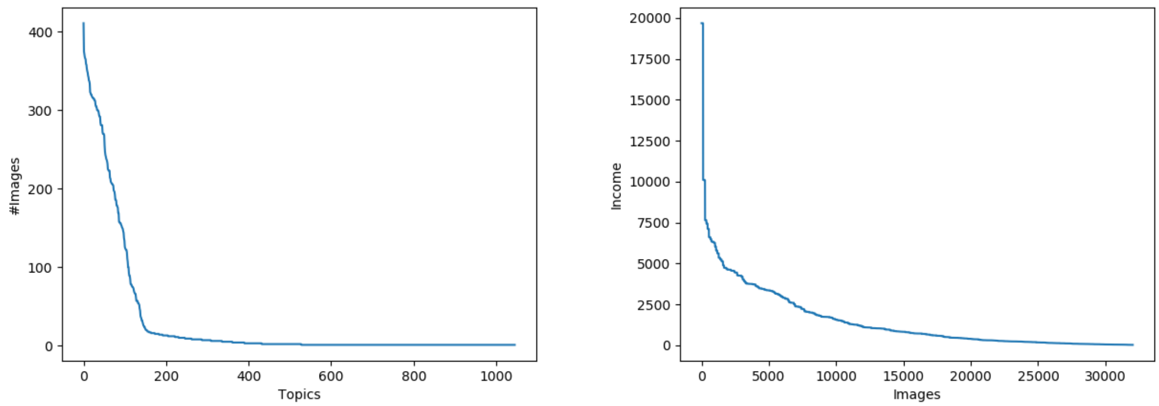
Users have tested Twitter's saliency cropping algorithm for racial and gender bias, which has been further studied by Yee et al. [2021] who have shown bias towards lighter-skinned individuals and women.

I raised the question whether there are further social groups which can be unintentionally harmed by this type of algorithm. One hypothesis was, that a certain type of attractiveness or wealth bias may be amplified on the platform. Dollar Street[1] is a project which collects photos from around the world and visualises them by the income of the family the object of the photo belongs to. Their dataset seemed optimal for testing this hypothesis. My expectation was that if there is any pattern, saliency cropping will be biased towards photos of higher income individuals. However, based on the initial results presented here, the picture is more complex than that. Saliency cropping seems to favour clutter, shiny, intricate things and logos. These features are present in photos from various parts of the income spectrum, however, the category or topic of the scenes do show a pattern of income bias.

I found that the algorithm is biased towards *cheap* rooms and spaces and towards *expensive* objects if the photo is zoomed in. This can potentially be problematic in various ways. It can reinforce stereotypes of how people live in different social classes. It can work as a form of erasure by exaggerating higher income class narratives. Lower income individuals' objects can be under-represented, whereas their cluttered rooms over-represented. This all can have an effect on people's reputation, by misrepresenting their homes and possessions.

In Section 2 I introduce the Dollar Street dataset I used. Section 3 contains the description of preparing the data and conducting the experiments. I present the results in Section 4 and estimate a grade in Section 5. I conclude and discuss future work in Sections 6.

[1] `https://www.gapminder.org/dollar-street`



(a) Number of images in each topic.

(b) Income labels for each image (USD/month/household).

Figure 1: Image and income distribution in Dollar Street.

2 Dollar Street Dataset

I used an image dataset labelled with incomes provided by Dollar Street (DS). DS is a project for demonstrating in what circumstances people live around the world with different financial backgrounds. A team of photographers have documented over 264 homes in 50 countries so far. In each home the photographer spends a day taking photos of up to 135 objects, like the family's toothbrushes or favorite pair of shoes. All photos are then tagged by household function, family name and income. They calculate the income of each by measuring consumption rather than salary. This means that if a family grows all the rice they eat each month, then the value of this rice will be included in the total consumption of the home[2] [3].

Currently, there are 32 034 images in the dataset belonging to 1 046 topics. All the data is publicly available on the DS website. The creators were so kind to share the data with me, including income labels in an easily processable format. The dataset contains a float income label for every image. Furthermore, each image has a topic label out of the 1 046, such as *Toothbrush*, *Home* or *Phone*.

Figure 1 includes the distribution of images across topics and the income distribution across all images in the dataset.

They shared this data under the conditions of me not re-sharing it. Because of GDPR, I cannot share examples including any parts of people. The data is publicly available for browsing purposes. In theory this work is reproducible, however, I suggest asking for the pre-processed data from Dollar Street.

3 Data Preparation and Experiments

I constrained the task of measuring bias for a continuous income scale to a binary task. I created a subset of the DS dataset with binary *cheap* / *expensive* labels. In order to make the binary labels reasonably separable I only used images with income labels from the two extreme ends of the income spectrum. In each topic I ordered the images by income then divided them to 4 parts with equal cardinality (quartiles). The lowest quartile then got the *cheap*, whereas the highest quartile received the *expensive* label. Images which fall in the middle two quartiles were not used for these experiments.

For testing the income bias of the image cropping algorithm I used the same methodology as Yee et al. [2021] for race and gender binary categories. Using their methods on each pair, I sample one image independently and uniformly at random from each of the two groups and attach those two images horizontally (padding black background when images have different heights). I ran the experiments for each topic separately. For the time constraints of this challenge I only

[2] `https://www.gapminder.org/dollar-street/about`

[3] `https://drive.google.com/drive/folders/0B9jWD65HiLUnRm5ZNWlMSU5GNEU?resourcekey=0-4rjWstzby3z96urmt8QgpA.`

Table 1: First 31 topics with the highest number of images

Topic	#Images
Home	411
Floor	375
Front door	371
Everyday shoes	368
Lock on front door	366
Wardrobe	365
Ceiling	358
Nicest shoes	353
Street view	350
Cooking pots	348
Wall inside	342
Toothbrush	338
Phone	334
Grains	323
Soap for hands and body	322
Dish washing soap	320
Kitchen sink	320
Wall decoration	317
Shampoo	316
Spices	316
Shower	315
Water outlet, drainage	315
Washing detergent	313
Books	312
Kitchen	307
Roof	305
Salt	305
Power outlet	303
Medication	300
Street detail	300
Tooth paste	300

had time to run the experiments for the first 31 topics with the highest number of images[4]. Although, as we can see in Figure 1a, the number of images drop rapidly until around the 150th topic (ordered by image count). Table 1 shows the 31 topics I used and the number of images they include.

I modified the code provided for this challenge[5]. The code for preparing DS data and performing the income bias experiments is publicly available on Github[6].

4 Results

Figure 2 represents the bias towards cheap objects on attached pairs of two photos, one from the *cheap* one from the *expensive* category. X axis shows $P(cheap) - 0.5$ for better visibility of the bias compared to the average. As we can see there are a few extreme topics with high *expensive* bias e.g. *Water outlet, drainage*; *Washing detergent*; *Everyday shoes* and a few with *cheap* bias e.g. *Ceiling*; *Floor*; *Wardrobe* and *Wall inside*. Figure 3 shows example bias probability scores and Empirical cumulative distribution functions (ECDFs) of maximum scores. Figures for all topics, and figures of mean and median ECDFs, are available in the project Github repo[7].

[4] I tried the first 36 but could not retrieve images for 5 of the topics: *Cups/mugs/glasses*, *Dish washing brush/cloth*, *Hair brush/comb*, *Stove/hob*, *Washing clothes/cleaning*.

[5] `https://github.com/twitter-research/image-crop-analysis`

[6] `https://github.com/anitavero/twitter_thumbnail`

[7] `https://github.com/anitavero/twitter_thumbnail/tree/main/data/results`

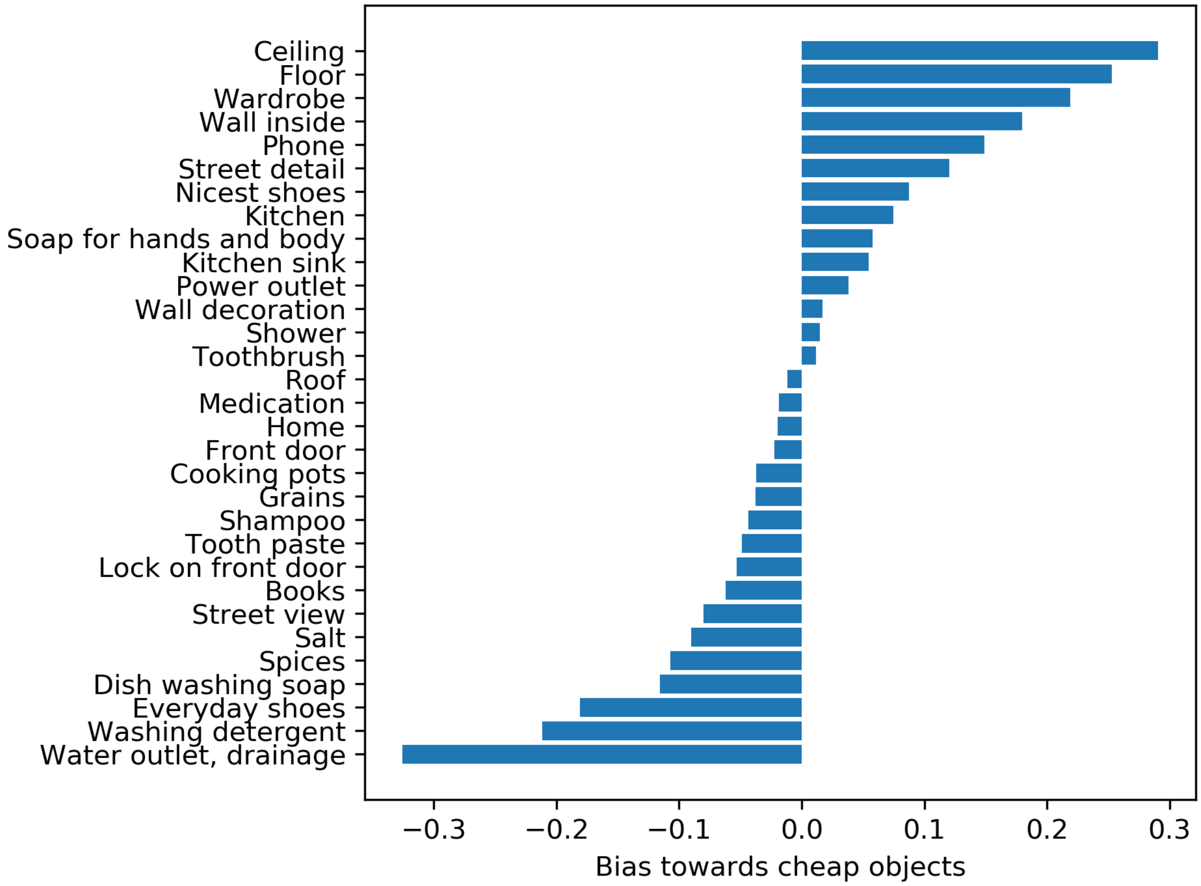


Figure 2: Bias towards *cheap* objects on attached *cheap*/*expensive* pairs of photos in each topic. The more positive the score is the bigger the bias towards *cheap*, the more negative it is the bigger the bias towards *expensive*.

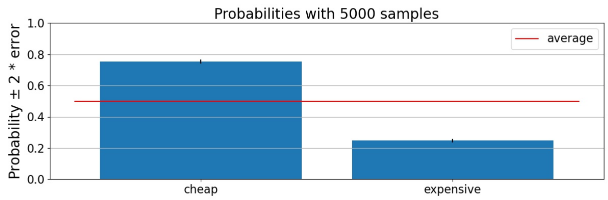
We can observe that topics related to rooms and spaces, such as ceiling, floor, wall tend to gain a *cheap* bias, whereas smaller objects, such as detergents, soaps, spices, books, tooth paste, shampoo receive an *expensive* bias from the cropping method. Based on qualitative evaluation of the saliency maps, I assume that the algorithm is drawn to clutter in broader spaces, such as walls, ceilings and floor. Furthermore, it attends to shinier objects, objects with logos on it as well as shelves full of items as opposed to half empty ones. Cluttered spaces tend to occur in lower income households, whereas shinier objects, fuller shelves and logos tend to appear in higher income households. Figure 4 includes example crops and saliency maps, where a *cheap* object or space is on the left, and an *expensive* is on the right. Figures 4a is an example for a salient logo, 4b and 4c showcase shelves full of objects. Figures 4d and 4e present broader spaces, where the low income one is more cluttered.

5 Self-Grading Assessment

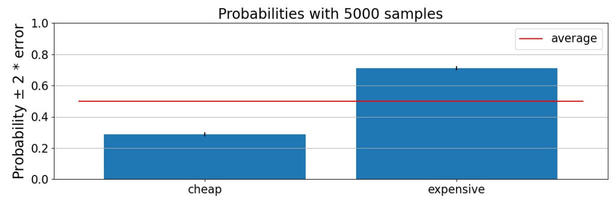
I demonstrate a risk of disproportional income bias on photos when the saliency algorithm is used to automatically crop images containing multiple objects or spaces from different ends of the income spectrum.

I would categorize this submission as unintentional harm. Below I summarise the reasons for choosing harm categories and estimate a Harm Base Score:

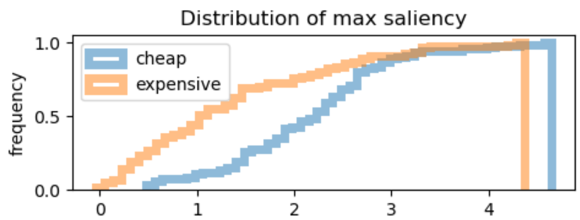
- Stereotyping: 20 points. Both rich and poor could be stereotyped by the focus on cluttered spaces or expensive objects.
- Erasure: 15 points. Higher income classes narrative can be exaggerated.



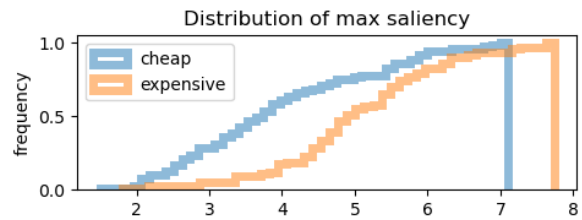
(a) Floor bias



(b) Washing detergent bias



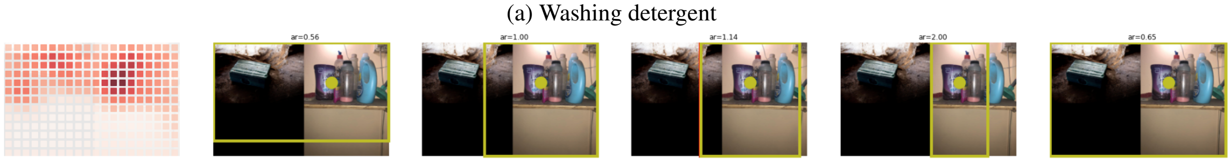
(c) Floor EDCF of maximum saliency scores.



(d) Wshing detergent EDCF of maximum saliency scores.

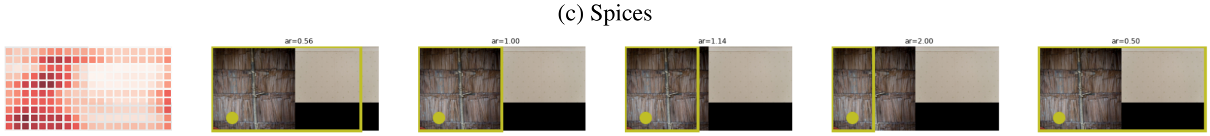
Figure 3: Examples for bias probabilities ((a) and (b)); and Empirical cumulative distribution functions (ECDFs) of maximum scores ((c) and (d)).

(a) Washing detergent

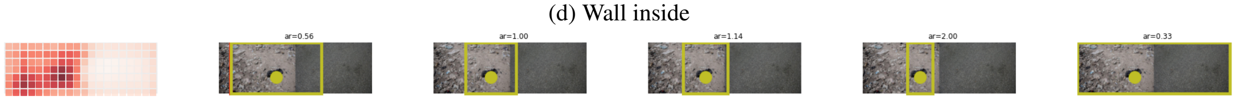


(b) Washing detergent

(c) Spices



(d) Wall inside



(e) Street detail

Figure 4: Example crops for rooms and objects. *cheap* object or space is on the left, and an *expensive* is on the right. (a), (b), (c) show objects, (d) and (e) include broader spaces.

- Under-representation: 20 points. Lower income people's objects can be under-represented, whereas their cluttered spaces can be over-represented.
- Reputational: 8 points. Lower income people's reputation can be harmed my misrepresenting their photos of their homes and possessions.

Harm Base Score: 20 + 15 + 20 + 8 = 63

Multiplier Factors:

- Damage: I showed that harm can be measured along the axix of inceom. (Damage = 1.2)
- Affected Users: It can affect all users. (Affective User Score = 1.3)
- Likelihood or Exploitability:
  - This harm is estimated to have occurred on Twitter daily (Likelihood = 1.3)
  - Submitted harm was classified as an unintentional harm therefore I receive no exploitability grade.
- Justification: The findings are promisin but preliminary. More experiments could be performed with more time, which are discussed in the Future Work 6. (Justification = 1.0)
- Clarity: The submission included detailed instructions and code that allowed any person to rproduce it. Limitations of the methodology and the data used in analysis were culturally situated and well documented. (Clarity = 1.5)
- Creativity: This submission includes a new dataset used in a completely new way. Therefore it can qualify for some extra points.

The overall score is: 63 base points x MF(1.2 + 1.3 + 1.3 + 1.0 + 1.5) = 396.9 + extra creativity points :)

6 Conclusion and Future Work

I raised the question whether income based class bias could be shown using Twitter's saliency cropping algorithm. I introduced a new way of using Dollar Street data in order to test that and found that the algorithm is biased towards *cheap* rooms and spaces and towards *expensive* objects.

For the limited time of the hackaton I could only test for 31 topics out of the 1 046, although they were the ones with the highest number of images and the image count drops around the 150th topic. Nevertheless, testing on the whole dataset would be the obvious next step. Furthermore, the current way of creating binary classes using quartiles could further be tuned, since across the dataset a heavy tailed income distribution shows with a few very high incomes. More fine grained discretisation of the income spectrum could be experimented with. Another interesting experiment would be to measure bias across all pairs of images in every topic, without any discretisation and visualise bias probabilities in a continuous fashion. Dollar Street has further information in the form of country labels. If I had had more time I would have used that too for testing potential country bias. For example one could compare countries on specific levels of income (e.g., USA vs. rest of the world) or use GDP based categorisation.

I hope the idea of using this type of dataset this way will inspire further research.

References

Kyra Yee, Uthaipon Tantipongpipat, and Shubhanshu Mishra. Image Cropping on Twitter: Fairness Metrics, their Limitations, and the Importance of Representation, Design, and Agency. *arXiv e-prints*, art. arXiv:2105.08667, May 2021.